



UNIVERSITÉ DE NANTES



IAE NANTES  
ÉCONOMIE & MANAGEMENT

ÉCONOMÉTRIE FINANCIÈRE HAUTE FRÉQUENCE

---

## Analyse de l'USD/CAD entre 1980 et 2013

---



Diane THIERRY

Enseignant : M. SÉVI  
Année universitaire : 2020-2021

Master 2 Économétrie et Statistiques, parcours Économétrie Appliquée

**Table des matières**

<b>I</b>	<b>Analyse économique</b>	<b>2</b>
I.1	Les données . . . . .	2
I.2	Propriétés de la série . . . . .	3
<b>II</b>	<b>Modélisations</b>	<b>5</b>
<b>III</b>	<b>Prévisions</b>	<b>7</b>

# I Analyse économique

Dans ce dossier nous allons étudier la série du taux de change des dollars américains et canadiens, notée **USD/CAD**. La valeur de clôture correspond à la valeur d'un dollar des États-Unis, exprimée en dollar canadien. L'USD/CAD fluctue avec les décisions monétaires prises par les banques centrales respectives, c'est pourquoi sa valeur peut être impactée par une modification des taux d'intérêt ou encore des programmes de rachat d'actifs effectués par l'une ou l'autre de ces grandes banques. La volatilité de cet indice est alors très forte avec une réaction des marchés appelée « effet d'annonce ».

Aussi, un des éléments les plus importants à prendre en compte dans l'analyse de l'indice USD/CAD est le **prix du pétrole**. Le Canada est effectivement l'un des grands pays producteurs de pétrole, et son économie est principalement basée sur l'exploitation de cette ressource naturelle. On observe historiquement une corrélation de l'ordre de 70 à 80% entre la valeur du dollar canadien et celle du pétrole. Lorsque les prix du baril progressent à la hausse, le CAD a donc tendance à s'apprécier lui aussi.<sup>1</sup>

Cette analyse vise à modéliser la **variance conditionnelle** des rendements de l'actif USD/CAD, ce qui est rendu possible par l'utilisation de modèles HAR c'est-à-dire *Heterogeneous Autoregressive Model*. Ils supposent que la volatilité réalisée dépend de sa valeur passée à un jour, de la moyenne des valeurs de la semaine passée et du mois passé, reflétant ainsi les différents composants du processus de volatilité.

Dans un premier temps nous ferons quelques statistiques descriptives dans le but de connaître la base et les différentes variables qu'elle contient, puis nous les modéliserons à 3 horizons distincts (une journée, une semaine et un mois ouvrés) et enfin nous essayerons de prévoir les variations futures à 1 journée.

## I.1 Les données

Les données à haute fréquence dont nous nous servirons dans l'analyse s'étendent du 21 juillet 1980 au 16 août 2013, pour un total de 6869 observations.

Nous disposons initialement de 6 variables :

- **date** : la date du jour au format DD/MM/YYYY
- **closing price** : la valeur du cours en fin de journée
- **return** : le rendement du cours comme le logarithme de la différence de l'indice entre 2 journées
- **RV** : la variance réalisée chaque jour avec un calcul toutes les 5 minutes
- **MedRV** : inclut dans la RV la variabilité du saut réalisé (permet notamment des prévisions plus précises)
- **Jump** : les sauts de la série, *i.e.* événements rares et importants qui impactent la volatilité

Au fur et à mesure de l'analyse nous serons amenés à construire de nouvelles variables à partir des 6 existantes, pour apporter des informations supplémentaires notamment pour l'estimation et la prévision de la volatilité réalisée.

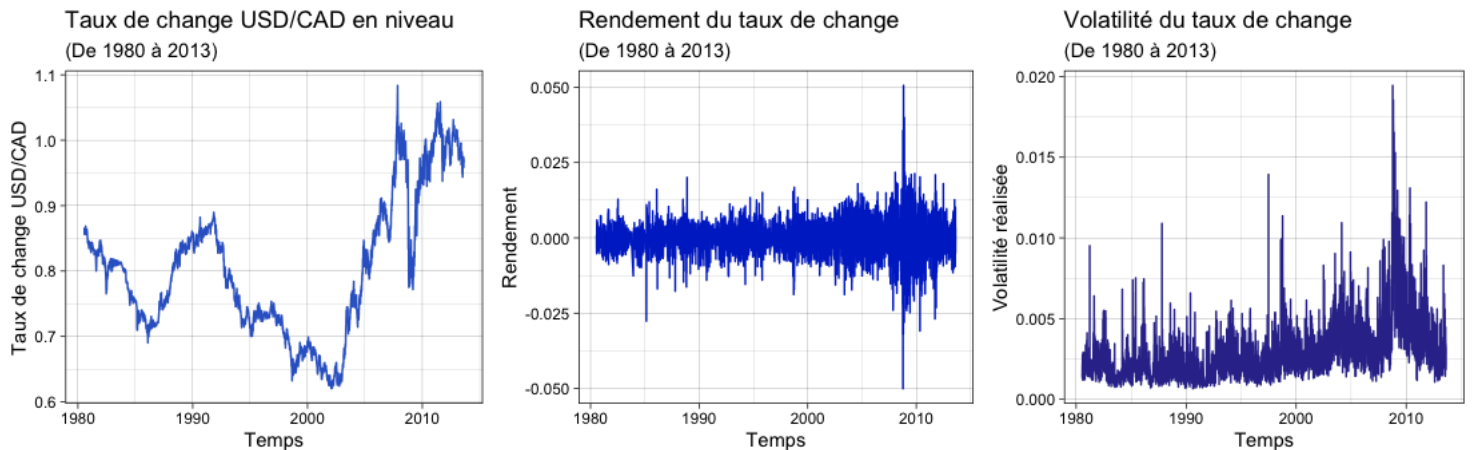
---

1. <https://www.strategie-bourse.com> (consulté le 1er février 2021)

## I.2 Propriétés de la série

Sur la figure n°1 nous pouvons observer les évolutions de la série USD/CAD en niveau, celles du rendement de l'actif ainsi que la volatilité réalisée que nous avons calculée en prenant la racine carrée de la variance réalisée.

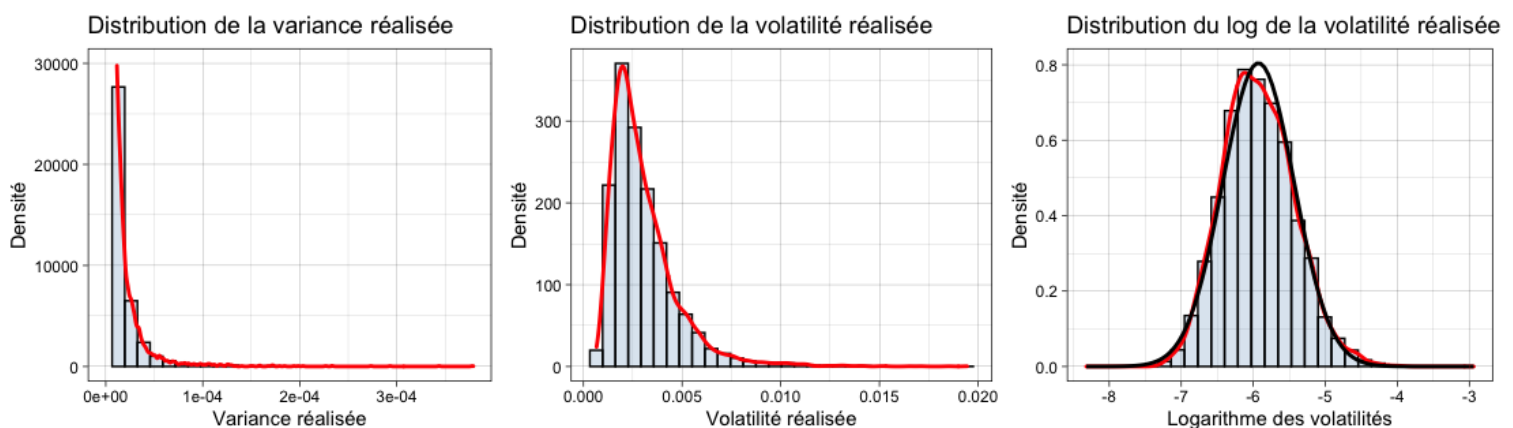
FIGURE 1 – Évolution de la série brute, de son rendement et de sa volatilité



Sur les 33 ans d'évolutions on peut distinguer pour le taux de change en niveau 4 principales phases : la première va de **1980 à 1986** où le taux passe de 0.86 à 0.6997 au plus bas le 3 mars 1986, ce qui traduit une dévaluation du dollar canadien puisqu'il faut alors plus d'argent pour obtenir un même dollar américain. Une deuxième période s'étend de **1986 à 1992** où le ratio revient à ses valeurs du début des années 80, puis chute à nouveau **jusqu'en 2002** où le taux de conversion n'est plus que de 0.63 au 5 août, c'est-à-dire qu'il faut presque \$1.6 canadien pour \$1 américain. **À partir de 2002**, le dollar canadien prend de la valeur et dépasse même la parfaite égalité USD/CAD=1 le 1er février 2012.

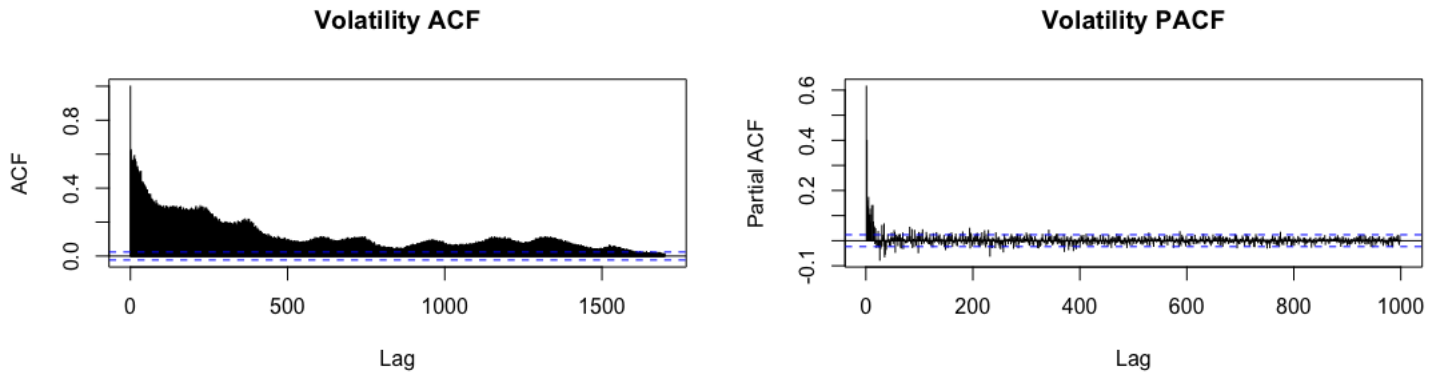
Les rendements de cette action fluctuent autour de 0 avec un rendement moyen de 0.001% ; on observe une plus grande volatilité à partir de 2008-2009 ce qui correspond aux répercussions de la crise économique et financière mondiale. Cela est largement confirmé par l'évolution de la volatilité du taux de change USD/CAD puisqu'à part quelques journées avec une forte volatilité de 1980 à 2008, c'est à partir de 2008 que les volatilités augmentent pour se stabiliser quelques mois plus tard.

FIGURE 2 – Distributions des valeurs des séries du taux de change



La figure n°2 montre que toutes les valeurs de la **variance réalisée** sont positives, avec une certaine inflation en 0 qui concerne près de 2800 observations. L'écart-type de cette série observable à travers la **volatilité réalisée** s'apparente davantage à une loi normale où les valeurs les plus fréquentes correspondent à une volatilité de 0.002. Enfin, lorsque l'on regarde l'histogramme de distribution du **logarithme de la volatilité réalisée**, on voit que celui-ci s'apparente très fortement à la loi normale signifiée par la courbe noire. L'hypothèse de normalité de la distribution est une hypothèse forte pour l'application des moindres carrés généralisés, c'est pourquoi nous nous attendons à de meilleurs résultats (qualité d'ajustement du modèle plus élevée) pour les estimations du logarithme de la volatilité.

FIGURE 3 – Fonctions d'autocorrélation de la variance réalisée



Nous constatons au travers de la figure n°3, que les observations sont fortement corrélées entre elles puisqu'il faut 1700 retards pour que le coefficient d'autocorrélation ne soit plus significatif sur l'ACF, et plus de 1000 retards sur la PACF. Il existe donc une forte dépendance et persistance des données dans le temps, ce qui justifie entre autres l'utilisation de modèles HAR introduits en section précédente.

Maintenant que nous avons observé nos séries, nous chercherons dans une seconde partie à modéliser puis prévoir la volatilité réalisée.

## II Modélisations

Dans cette partie nous cherchons à modéliser la variance et la volatilité réalisées, ainsi que la volatilité sur 3 horizons différents : une journée, une semaine financière soit 5 jours et un mois financier soit 22 jours ouvrés. Nous obtiendrons donc 9 modèles, où chaque série est modélisée avec ses valeurs passées. Ainsi, on explique la variance réalisée chaque jour par la variance du jour passé, la moyenne de la semaine et du mois passés, ainsi que du saut retardé - comme expliqué précédemment. Nous avons créé sous Excel ces différentes variables et avons importé la base directement sous R. Pour l'application de ces modèles HAR nous n'avons pas traité les données atypiques puisqu'ils y sont peu sensibles, du fait que les données importantes sont "écrasées" par les modèles.

TABLE 1 – Estimations HAR quotidiennes, hebdomadaires et mensuelles

$$RV_{t,t+h} = \beta_0 + \beta_D RV_t + \beta_W RV_{t-5,t} + \beta_M RV_{t-22,t} + \beta_J J_t + \varepsilon_{t,t+h}$$

$$(RV_{t,t+h})^{1/2} = \beta_0 + \beta_D RV_t^{1/2} + \beta_W RV_{t-5,t}^{1/2} + \beta_M RV_{t-22,t}^{1/2} + \beta_J J_t^{1/2} + \varepsilon_{t,t+h}$$

$$\log(RV_{t,t+h}) = \beta_0 + \beta_D \log(RV_t) + \beta_W \log(RV_{t-5,t}) + \beta_M \log(RV_{t-22,t}) + \beta_J \log(J_t + 1) + \varepsilon_{t,t+h}$$

$h$	$RV_{t,t+h}$			$(RV_{t,t+h})^{1/2}$			$\log(RV_{t,t+h})$		
	1	5	22	1	5	22	1	5	22
$\beta_0$	6.264e-07** (1.986e-07)	8.204e-07*** (1.191e-07)	1.525e-06*** (1.036e-07)	1.388e-04*** (3.058e-05)	1.794e-04*** (1.907e-05)	3.117e-04*** (1.696e-05)	-0.120*** (0.024)	-0.167*** (0.015)	-0.289*** (0.014)
$\beta_D$	0.116*** (0.014)	0.082*** (0.008)	0.047*** (0.007)	0.155*** (0.015)	0.102*** (0.009)	0.061*** (0.008)	0.172*** (0.014)	0.112*** (0.009)	0.067*** (0.008)
$\beta_W$	0.189*** (0.030)	0.098*** (0.018)	0.323*** (0.015)	0.295*** (0.023)	0.238*** (0.018)	0.292*** (0.016)	0.361*** (0.027)	0.309*** (0.017)	0.259*** (0.016)
$\beta_M$	0.642*** (0.028)	0.753*** (0.017)	0.506*** (0.002)	0.508*** (0.026)	0.605*** (0.016)	0.547*** (0.014)	0.421*** (0.024)	0.513*** (0.016)	0.561*** (0.014)
$\beta_J$	8.551e-04*** (1.942e-04)	-1.952e-04 (1.165e-04)	-1.361e-04 (1.013e-04)	-3.311e-03** (1.008e-03)	-2.571e-03*** (6.287e-04)	-2.023e-03*** (5.591e-04)	-10.388* (4.767)	-10.654*** (3.096)	-5.946* (2.780)
$R^2_{HAR-RV-J}$	0.5403	0.7571	0.7871	0.6246	0.8036	0.8249	0.6207	0.7853	0.8017
$R^2_{HAR}$	0.5391	0.757	0.787	0.6241	0.8031	0.8246	0.6205	0.785	0.8016

La table 1 nous montre ces 9 modélisations HAR appliquées sur les données du taux de change USD/CAD, avec les coefficients estimés et la qualité du modèle avec et sans sauts.

- En règle générale on voit que la qualité du modèle est plus forte lorsque l'on modélise la variance ou la volatilité **mensuelles**, plutôt que quotidiennes ou hebdomadaires ; on peut donc conclure que ce type de modèle est dans notre cas meilleur pour expliquer la variance conditionnelle de **long terme**. Si l'on regarde encore la qualité d'ajustement du modèle, on voit qu'elle est aussi plus élevée lors des modélisations de la volatilité réalisée et de son logarithme, par rapport à la RV (1ère équation). En moyenne sur les 3 horizons, le  $R^2$  diffère de 5.62% entre les estimations de la variance et celles de la volatilité réalisée. Par ailleurs on ne note pas de différence significative dans cette mesure entre les modélisations de la volatilité réalisée et de son logarithme.

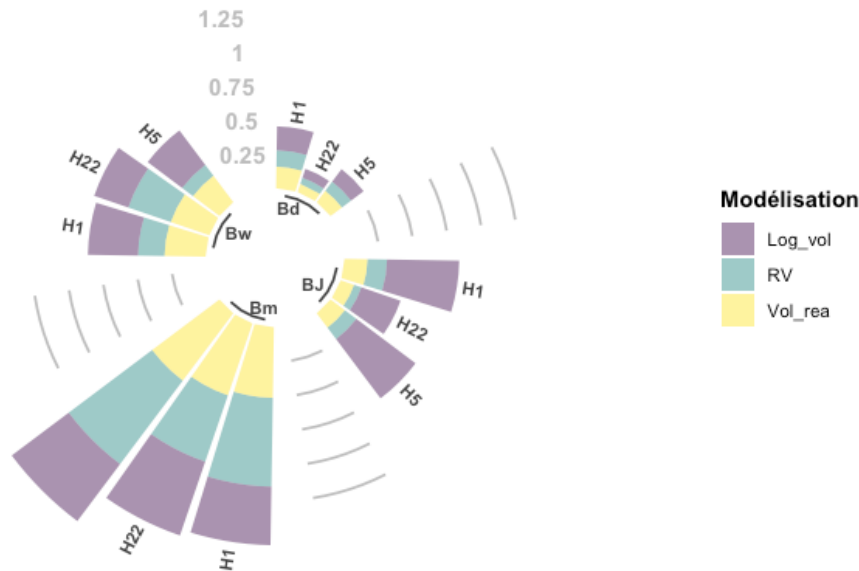
- Nous voyons aussi que la majorité des coefficients sont **significatifs** même au seuil de risque de 0.1%, excepté les sauts sur lesquels nous reviendrons prochainement.

- Un autre constat valable pour la modélisation des 3 séries est que les coefficients estimés du jour passé  $\beta_D$  sont plus importants aux horizons d'une journée, puis diminuent jusqu'aux horizons d'un mois. Cela semble logique, et nous en déduisons que pour expliquer la volatilité du cours USD/CAD à

une journée, c'est sa valeur de la veille qui impactera le plus le taux aujourd'hui. Les constats ne sont pas aussi tranchés sur l'impact des variables explicatives sur  $Y$  à horizon d'une semaine et d'un mois, cependant on peut constater que les coefficients  $\beta_D$ ,  $\beta_W$  et  $\beta_M$  sont tous positifs pour les 9 modèles. Ceci confirme la forte persistance mise en avant par les ACF et PACF, puisque le coefficient en rouge montre par exemple que lorsque la volatilité réalisée de la semaine passée augmente d'une unité, alors celle du mois à venir augmentera elle aussi de 0.605 point.

- Concernant les **sauts** on voit qu'au seuil de risque de 5%, ils sont tous significatifs excepté pour expliquer la variance réalisée de la semaine et du mois à venir. Concrètement, cela montre que les événements rares qui impactent les séries ont un plus grand effet sur la volatilité de la série que sur sa variance, cet effet est négatif. Par ailleurs, étant significatifs pour la deuxième et la troisième équation, les sauts permettent d'améliorer l'ajustement du modèle mais très légèrement puisqu'en moyenne sur ces 6 modèles, la différence dans le  $R^2$  avec et sans *jump* est de 0.03%. Finalement, on voit que l'effet quantifié par le coefficient estimé du  $\beta_J$  pour le logarithme de la volatilité réalisée est très fort puisque les coefficients sont de -10.38, -10.65 et -5.95 aux horizons quotidiens, hebdomadaires et mensuels. La survenance d'événements rares impacte donc davantage la volatilité que la variance réalisée.

FIGURE 4 – Barplot circulaire des valeurs des coefficients des 9 modèles



Enfin, la figure n°4 ci-dessus permet de constater visuellement les valeurs des coefficients estimés par les 9 modèles, selon la série modélisée et l'horizon d'estimation. Les coefficients des sauts étant négatifs pour certains, très faibles pour d'autres ou encore très élevés, nous avons procédé à un réajustement afin de comparer non pas la valeur exacte mais les ordres de grandeur entre chaque série. Nous avons donc pris la valeur absolue de  $\beta_J * 200$  pour la RV,  $*50$  pour la volatilité réalisée et  $/20$  pour le logarithme de la volatilité réalisée.

On voit d'emblée que les coefficients (non modifiés) les plus grands sont ceux de l'effet mensuel  $\beta_M$  plus particulièrement lorsque l'on cherche à expliquer les valeurs de la semaine à venir ( $H=5$ ). De même, comme évoqué précédemment, les sauts sont plus importants dans l'explication du logarithme de la volatilité que des 2 autres séries. On peut aussi confirmer que les valeurs des coefficients du jour passé ont un plus grand effet lorsque l'on cherche à modéliser les valeurs d'aujourd'hui.

Maintenant que nous avons modélisé la variance et la volatilité réalisées ainsi que le logarithme de la volatilité à 3 horizons différents, nous allons essayer de prévoir au mieux ces 3 séries jour par jour.

### III Prévisions

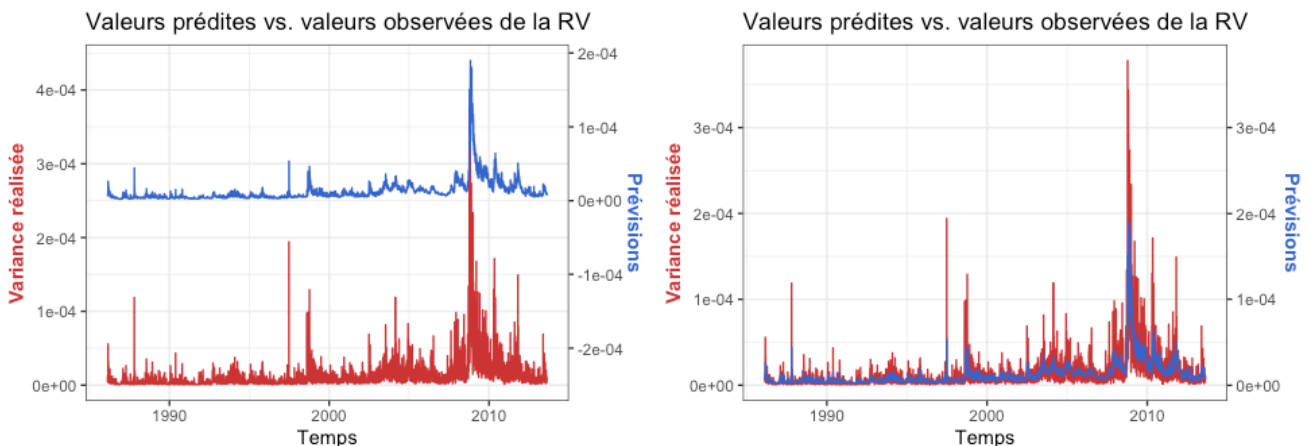
Les prévisions que nous réaliserons se feront pas à pas avec une fenêtre roulante de 1000 observations ; on suppose pour cela qu'avec les données disponibles jusqu'en  $t-1$  nous sommes en mesure de prévoir les valeurs en  $t$ . Nous construisons une boucle sous R dans laquelle nous modélisons sur 1000 observations, puis nous prédisons la 1001<sup>è</sup> que nous comparons à la valeur réelle pour obtenir une erreur, etcétera en décalant d'une observation à chaque itération de la boucle. Le code sous R est visible ci-dessous :

FIGURE 5 – Boucle de prévisions sous R

```
h=6847-1000
i=1
prev <- matrix(nrow=h, ncol=1)
erreurs <- matrix(nrow=h, ncol=1)
for(i in 1:h)
{
  fit <- lm(RV~RVd+RVw+RVm+`lag(Jump)`, data=df[(23-1+i):(1022-1+i),]) #modèle
  prev[i,1] = predict.lm(fit, newdata=df[(1022+i),]) #prévisions
  erreurs[i,1] <- df$RV[1022+i]-prev[i,1] #erreur de prev
  i=i+1
}
erreurs
Resultats_RV <- df[1023:6869,c(1,4)]
Resultats_RV$Prev <- prev
Resultats_RV$erreurs <- erreurs
```

Nous obtenons ainsi 5847 prédictions allant du 18 février 1986 au 16 août 2013 (fin de la base). Nous stockons le tout dans un data frame avec les dates, les valeurs observées, les valeurs prédites et les erreurs (simple différence de ces deux dernières), et obtenons ainsi 3 bases ; pour la variance réalisée, pour la volatilité réalisée et pour son logarithme.

FIGURE 6 – Prévisions et valeurs réelles de la variance réalisée

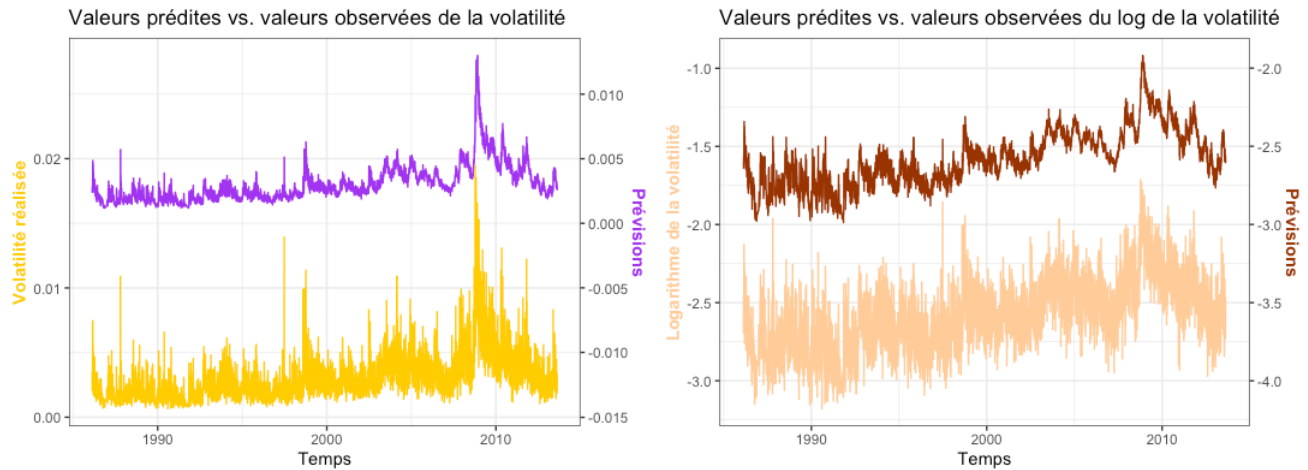


En figure n°6 nous pouvons observer les **prévisions** et les valeurs réelles, d'une part mises à échelles différentes, et d'autre part superposées pour pouvoir mieux observer les points communs et les différences. On voit que les prévisions faites à partir du modèle *HAR* sont extrêmement fiables puisqu'elles suivent de près les valeurs réelles de la variance réalisée, comme visible sur les graphiques. En revanche, on voit que les prévisions sont beaucoup moins volatiles que la RV observée ; ainsi



on retrouve les mêmes évolutions mais avec une courbe moins épaisse. En termes de prévisions à 1 jour de la variance réalisée, les modèles HAR anticipent mal les **grandes variations** mais anticipent correctement la **tendance** des évolutions. Voyons ce qu'il en est à présent des prévisions des volatilités.

FIGURE 7 – Prévisions et valeurs réelles de la volatilité réalisée et de son logarithme



Nous observons la même chose que pour la variance réalisée à savoir que les prévisions réalisées à partir des modèles HAR suivent les mêmes évolutions : on retrouve les mêmes pics et tendances à la hausse ou à la baisse, mais on observe tout de même des variations plus faibles dans les prédictions que dans les valeurs observées. Il apparaît alors que les modèles HAR prédisent moins bien les grandes volatilités.

En termes d'**erreur de prévisions**, le tableau n°2 nous informe que c'est le logarithme de la volatilité réalisée qui est le plus mal prédit puisque les erreurs moyennes au carré atteignent 0.00124 contre 1.8144e-10 et 2.0635e-11 respectivement pour la variance et la volatilité réalisées. Dans l'ordre des erreurs, on voit que c'est la variance réalisée quotidienne qui est la mieux prédite par les modèles puisque pour les 4 types erreur c'est le taux le plus faible.

TABLE 2 – Erreurs associées aux différentes prévisions

	$RV_{t,t+h}$	$(RV_{t,t+h})^{1/2}$	$\log(RV_{t,t+h})$
<b>MSE</b>	1.8144e-10	2.0635e-11	0.00124
<b>MAE</b>	6.3657e-06	1.2025e-06	0.01778
<b>MDSE</b>	1.0136e-11	9.2411e-14	5.3458e-05
<b>MDAE</b>	3.1836e-06	3.0399e-07	0.00731

Ainsi, les conclusions diffèrent des hypothèses émises au cours de l'analyse : les *Heterogeneous Autoregressive Model* prédisent mieux la variance réalisée et moins bien le logarithme de la volatilité réalisée, alors que lors de l'estimation, la qualité d'ajustement pour cette dernière série était l'une des plus fortes ( $R^2_{Log\_vol}=0.6207$  et  $R^2_{RV}=0.5403$  pour  $h=1$ ). Toutefois, la modélisation et la prévision de la variance conditionnelle du taux de change USD/CAD est satisfaisante, car nous obtenons des prévisions qui suivent les évolutions réelles. De plus, nous expliquons entre 54 et 82% de la volatilité tous horizons confondus.